

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/22426>

Please be advised that this information was generated on 2018-07-07 and may be subject to change.

Eur J Clin Chem Clin Biochem
1995; 33:157–169

© 1995 Walter de Gruyter & Co.
Berlin · New York

Desirable Routine Analytical Goals for Quantities Assayed in Serum

Discussion paper from the members of the External Quality Assessment (EQA)
Working Group A¹⁾ on analytical goals in laboratory medicine

By Dietmar Stöckl¹, Henk Baadenhuijsen², Callum G. Fraser³, Jean-Claude Libeer⁴, Per Hyloft Petersen⁵ and Carmen Ricós⁶

¹ *Institut für Standardisierung und Dokumentation im medizinischen Laboratorium e. V. (INSTAND e. V.), Düsseldorf, Germany*

² *Academisch Ziekenhuis Sint Radboud, Nijmegen, The Netherlands*

³ *Department of Biomedical Medicine, Ninewells Hospital and Medical School, Dundee, Scotland, U. K.*

⁴ *Instituut voor Hygiëne en Epidemiologie, Klinische Biologie, Brussels, Belgium*

⁵ *Klinisk Kemisk Afdeling (Centrallaboratoriet), Odense Universitetshospital, Odense, Denmark*

⁶ *Servicio de Bioquímica, Hospital General Universitario Vall d'Hebron, Po Vall d'Hebron, Barcelona, Spain*

(Received May 9/December 12, 1994)

Summary: The aim of the Working Group was to describe guidelines for deriving desirable analytical goals in laboratory medicine. First, a literature review is given of the different approaches used until now, and some of the most important studies are presented in detail. These approaches are then discussed critically, and the analytical goals proposed by the group are outlined with respect to monitoring and diagnostic testing. The group recommends that, most realistically, analytical quality specifications be biologically based. For diagnostic testing, the aim is achievement of accuracy, allowing the use of common reference intervals when populations are homogeneous for a given quantity. For monitoring (within an individual laboratory and performed with the same instrument), analytical performance should aim at stable operation and low imprecision compared with the within-subject biological variation. Method accuracy is also very important for the comparability of results from different laboratories or instruments.

Foreword

The aim of the Working Group was to describe guidelines for deriving desirable analytical goals in laboratory medicine. Desirable means that analyses of the quality demanded by these goals have a negligible analytical uncertainty (due to method bias, unspecificity, or imprecision) for a quantity present in a specimen. The attainment of such goals would increase measurement comparability between different methods and laboratories, allowing the use of common reference intervals and

patient treatment strategies. In view of the complex and partly controversial nature of this topic, the literature describing the different approaches used so far is reviewed, and some of the most important studies are discussed in detail [for recent reviews see also: *Hyltoft Petersen & Hørdor* (1) and *Fraser & Hyltoft Petersen* (2)]. It should be noted that in this review part, we cite the nomenclature used by the respective authors which, in turn, leads to inconsistencies in wording and terminology. During development of the Working Group's concept, we soon realized that no universally applicable approach could be established. Every application requires special consideration. Analytical quality specifications for monitoring may be different from those for diagnostic testing. The situation in emergency analysis may differ from that in routine analysis (other techniques might

¹⁾ The Working Group is one of four ad hoc groups meeting under the auspices of the European external quality assessment (EQA) organizers group. The Working Groups were initiated by A. Uldall following a meeting of EQA organizers and interested individuals in Cracow, Poland, in 1991 at Eurolab '91.

be preferable in the former due to the necessity for a short response time). Special population screening programmes may require detailed studies on the relation of cost to analytical performance. Specific clinical strategies require discussions with the clinicians involved. Therefore, the Working Group restricted itself to outlining general analytical goals for monitoring and diagnostic testing. These recommendations are presented in detail here. The setting of analytical goals for clinical strategies is, however, not discussed [for more information see *l. c.* (1)].

Introduction

Analyses in laboratory medicine are only useful if measurements of quantities allow the evaluation of a person's state of health. This means that diagnosis or exclusion of a disease, detection of a person's predisposition to a disease, or monitoring during medical treatment is based on comparison of a value with a set of appropriate population based reference intervals or with values previously observed from the same individual. Two general aspects have to be considered when any kind of measurement is performed: a) the "medical requirements" (clinical usefulness), i. e. what quality is needed for the measurement from the medical point of view, and b) the "analytical performance", i. e. how this desirable analytical quality can be specified, achieved, and controlled. Factors connected with medical requirements are: within- and between-subject biological variations of a quantity, distribution of analytical results between sick and healthy subpopulations, prevalence of disease, use of an analysis (e.g. single point testing, patient follow up, population screening), and time of analysis. The distribution of analytical results between sick and healthy subpopulations and the prevalence of a disease determine the "a priori" relevance of an analytical test used to decide whether a particular disease is present or not. A widely used approach to assess the relevance of a clinical test is the predictive value model which defines the diagnostic performance of a laboratory test by its diagnostic sensitivity, diagnostic specificity, predictive value, and efficiency. This model and other approaches are discussed in detail in recent articles by *Linnet* (3), *Büttner* (4), *Christensen* (5), and *Henderson* (6). Since they are not directly related to analytical goal setting, they will not be outlined here. Analytical performance is associated with: bias and imprecision of an analytical procedure, analytical specificity and detection limit, analysis costs, internal and external control, etc. Analytical performance and medical requirements both influence each other. Medical knowledge can guide analytical strategies, whereas analytical progress can promote better understanding of biological processes. Therefore,

contributions to goal setting for quantitative measurements in laboratory medicine reflect both analytical and medical points of view. Strategies for defining maximum analytical bias and imprecision include "state of the art" of performance (7), opinions of experts (8, 9), national recommendations (10, 11), needs for quality control of analytical processes (12–16), proposals derived from the viewpoint of clinicians (17–22), analysis of the effect of performance on interpretation in specific clinical situations (1, 23–26), and derivation of criteria based on intra- or inter-individual biological variation (27–37).

Description of the Different Strategies

Analytical performance and "the state of the art"

According to some authors, the current state of the art achieved in laboratory medicine is already sufficient (7). Although this argument may be correct in many situations, it is untrue in many other situations, e. g. in certain fields of immunochemical determinations, in the case of recently developed analytical techniques without thorough validation, or in cases where inherently superior methods seem to disappear from the market because faster and cheaper assays become available. This means that the state of the art is not stable: it can change for the better but it can also deteriorate. Therefore, clear criteria should be defined concerning the analytical quality that is desired in laboratory medicine. Where this is impossible, at least criteria should be elaborated which can aid in the distinction between good and poor methods.

Analytical goals advocated by experts or expert groups

Analytical goals advocated by experts (8, 9) are generally empirical in nature. They mostly reflect the experience gained in the special field in which the experts are involved. They often include compromises taking the state of the art very much into account and are not real goals for desirable analytical quality. The latter is especially true for the goals used by organizations responsible for external quality assessment or proficiency testing (10, 11, 38).

Analytical goals and the needs of quality control

Discussion of analytical goals has become more and more influenced by the need for the quality control of analytical processes. Some general comments on quality

control in laboratory medicine can be found in a series of recommendations of the International Federation for Clinical Chemistry (IFCC) (39, 40), and in an article by *de Verdier* et al. (41). The general statement from the viewpoint of quality control is that the better methods are, the less effort and money has to be invested in their control. On this basis, analytical imprecision expressed as standard deviation (s_a , see Annex for a list of abbreviations) of 1/4 of the total allowable error is recommended for internal quality control (14), which is half that originally proposed by the same group (42). Results exceeding the total allowable error indicate that the analytical method is out of control. Usually no estimates for the total allowable error (TAE) are derived in such works but the currently used strategies are assessed as to whether they fulfil the needs for internal quality control (12, 13). In addition, no theoretical goals are derived for method imprecision, but the state of the art is often used as a basis for discussion, in order to estimate the s_a /TAE relation necessary for appropriate internal quality control. The benefit of these studies is that they provide a variety of statistical methods for evaluating method performance in terms of bias and imprecision (42–44), they discuss the quality control procedures necessary for specific analytical performance (15, 16), and they compare these strategies with those used in other scientific fields, e.g. industrial process management (14).

Analytical goals derived from the viewpoint of clinicians

As discussed in the introduction, different types of medical decisions may be based on the results of analyses done in laboratory medicine. Therefore, the necessary quality of an analytical result depends both on the actual clinical situation and on the physician who requested the analysis. In some studies, attempts have therefore been made to derive analytical goals from the viewpoint of clinicians by sending questionnaires to experienced physicians in many different fields of laboratory medicine (17–22). They were asked for decision levels for common quantities in particular clinical situations or for changes which would cause their medical action (medical significant differences = Δ_{med}). In the work of *Skendzel* et al. (20), medically useful analytical coefficients of variation (CV_a 's) were calculated in the following way for monitoring: $CV_a = \Delta_{\text{med}}/[1.65 \times \sqrt{2}]$. When a patient's value was compared with a reference value, the calculation was as follows: $CV_a = \Delta_{\text{med}}/1.65$. Compared with other proposals, these coefficients of variation were often quite large and were the subject of severe criticism (45). On the other hand, it was stated (13) that the medically significant differences reported

in the work of *Skendzel* et al. (20) were very useful, provided that these values were interpreted so that they reflect the sum of preanalytical (s_p), within-subject biological (s_i), and analytical standard deviation (s_a): $s_t = [s_p^2 + s_i^2 + s_a^2]^{1/2}$. If values for maximum analytical coefficients of variation were then calculated from the medically significant differences, they would be much lower than those proposed by *Skendzel* et al. (20). On the other hand, it is possible that the medical decision limits themselves are based on the experience of the physicians with the performance of the analytical tests, and so more or less reflect the analytical state of the art. In other words, it may not be possible to derive analytical goals from medical decision limits because they themselves may be strongly influenced by analytical performance.

Analytical goals derived from specific clinical situations

Every clinical situation requires a specific analytical approach. Therefore, it is very difficult to establish a uniform concept for the estimation of allowable imprecision and bias. Examples of specific situations are: determination of haemoglobin A_{1c} for long term diabetic control (25), determination of theophylline in serum for therapeutic drug monitoring (26), determination of creatine kinase isoenzyme in the diagnosis of acute myocardial infarction, determination of blood thyroid-stimulating hormone in screening for congenital hypothyroidism, and monitoring cholesterol as a measure for the risk of coronary heart disease (23). Further applications can be found in a recent review by *Hyltoft Petersen & Hørdér* (1). In a recent article, *Schechtman & Sasse* (46) investigated the influence of analytical imprecision on the medical usefulness of lipid measurements. The interpretation of the analytical performance was based on the use of clinical cut-off points or comparison with desired concentrations. All of these approaches have in common that they evaluate the clinical consequences caused by the analytical imprecision and bias either together or independently. Many different statistical approaches are used to calculate the effects at different levels of complexity: e.g. sum of misclassifications, the number of false negative results, or economical optimization. These approaches usually do not generate general figures for analytical bias and imprecision, but they present models for optimizing the analytical performance for the specific clinical situations. Therefore, they are more or less restricted to specialists in the field. Nevertheless, the statistical background outlined in these studies can also be used for the optimization of analytical methods in other applications.

A similar approach, which mainly discusses overlap of healthy and sick subpopulations and prevalence of diseases in the context of analytical needs is shown by *Lott et al.* (47). They discuss the effect of analytical imprecision on the percentage of correct diagnoses at different levels of prevalence and population overlap in three different applications, namely, population screening, long term, and short term trend detection.

A general theory for the allowable analytical coefficient of variation in situations that specifically involve the monitoring of individuals was proposed recently (24). The following formula was derived: $CV_a \leq [(\Delta_c^2/2Z^2) - CV_i^2]^{1/2}$, where Δ_c is the percentage of change with clinical significance in monitoring serial results from an individual, Z is a statistical value for the probability of significance, and CV_i is the within-subject biological coefficient of variation. A critical point of this approach is the estimation of the Δ_c value. In the work mentioned above, Δ_c was often derived from the viewpoint of clinicians (questionnaires), which perhaps does not reflect the desirable analytical quality but might be much influenced by the analytical state of the art as mentioned above.

Analytical goals based on within- and between-subject biological variation

The approach of relating analytical performance to the biological variation was very much inspired by the ideas of *Tonks* (27). He empirically stated that the total allowable error (TAE) should not exceed one quarter of the reference interval (R), using the following formula: $TAE = \pm 1/4[\text{of the reference interval}]/[\text{mean of the reference interval}] \times 100\%$.

Cotlove et al. (28) proposed that analytical goals be based not on the reference interval, because this reflects both the analytical and the biological variation (48), but on the composite biological standard deviation (s_c), which includes the within- (s_i) and the between-subject or group standard deviation (s_g). They stated that the analytical standard deviation should not exceed 0.5 times the composite biological standard deviation: $s_a \leq 0.5 s_c$, with $s_c = [s_i^2 + s_g^2]^{1/2}$, because then the analytical imprecision adds only 12% to the total test result variability: $s_t = [(0.5 s_c)^2 + s_c^2]^{1/2} = 1.12 s_c$. In fact, this formulation is similar to the one originally proposed by *Tonks* (27) because the total allowable error defined by him was meant as $2 s_a$. In addition, 95% of the reference interval represents the mean $\pm 2 s_c = 4 s_c$, if corrected for the analytical imprecision. *Tonks'* formula can then be written as: $2 s_a \leq 0.25 \times 4 \times s_c$, which is the same as the formula derived by *Cotlove et al.*

Following this hypothesis, data on biological variation were collected and goals for a variety of quantities were agreed on at two international conferences (30, 31). This data base was expanded and collected in a recent review (49), although data on individual within-subject biological variations have only been described for a few quantities (33, 50). It was further recommended (31) that the composite biological variation be used for group screening, and the within-subject biological variation for single individual testing.

The concepts outlined above mainly deal with method imprecision, but it has become increasingly apparent that method bias (expressed as the deviation from the "true value") should be included when analytical goals in laboratory medicine are derived. Therefore, *Harris* (51) expanded his original work (32) so that bias (B) and imprecision were both taken into account. The original equation: $s_a \leq 0.5 s_i(s_c)$ was thus changed into $[s_a^2 + B^2]^{1/2} \leq 0.5 s_i(s_c)$. In addition, the author focused in this work on diagnostic testing, especially when comparing patients' data with a reference interval of the healthy subpopulation, an idea originally proposed by *Tonks* (27) and later by *Glick* (29) [a series of recommendations on the theory of reference values was recently published by the IFCC, see l. c. (52) and articles cited therein]. As stated before, the reference interval reflects the within- and between-subject biological variation and, if the 95% interval is chosen, then $R = 4 s_c$. The formula for the allowable standard deviation (taking into account method bias) of an analytical method for single diagnostic testing becomes: $s_a \leq R[1/80 - 4/5 \times B^2/R^2]^{1/2}$, or $B \leq 0.125 R$, if s_a is zero. As a rule of thumb, he still proposed $s_a \leq 0.5 s_i$ for single serial testing, and $s_a \leq 0.1 R$ for single diagnostic testing in the absence of method bias.

Ross (53) uses the preanalytical test diagnostic efficiency as a starting point. He considers in particular the case in which the best performance is needed. This is the case when the means of the healthy and sick subpopulations are only 1.88 SD's apart. The preanalytical diagnostic efficiency is then 82.5%, which means that 82.5% true diagnostic statements can be made in the ideal case. In that situation, he examines the loss of diagnostic efficiency depending on the bias or the imprecision of the analytical method used, the test population variability and the probability for false rejection. The output of the model is the limit for allowable analytical error of a single result for a quantity in the clinical population under study. The advantage of this model is that it not only relates method bias and imprecision to the biological variance, but it additionally shows how both variables affect the confidence level of medical decisions. The disadvantage is its limitation to the situation

of a bimodal distribution with a 50% prevalence where the means are 1.88 SD's apart from each other. For individual single testing he proposes $s_a \leq 0.64 s_i$. In a recent article (54), this concept was expanded and $s_a \leq 0.5 s_i$ and $B \leq 0.25$ to $0.33 s_i$ were advocated. Further, total allowable analytical error should not exceed 1.25 to 1.40 times s_i .

Gowans et al. (35) based goals on the recommendations of the IFCC regarding reference intervals (52). They derived functions which allow the investigation of the influence of analytical bias and analytical imprecision (separately and in combination) on the percentage of the population outside reference limits. For a population sample size of 120, the IFCC stated that the maximum acceptable percentage outside each reference limit should be 4.6%. From this, a maximum analytical standard deviation (with no bias) of $s_a \leq 0.58 s_c$, and a maximum bias (with no imprecision) of $B \leq 0.25 s_c$ is calculated.

A similar approach was proposed by Stamm (55). For diagnostic testing he proposed a maximum method bias of $0.33 s_g$ together with a maximum standard deviation of $0.33 s_g$, in order to keep the sum of false classifications below 5% on each side of the reference interval. For monitoring, on the other hand, only limits for method imprecision were proposed. Further, these limits were not directly related to the biological variation, but were expressed as 1/4 of a predefined medical decision limit.

Klee (56) investigated the influence of imprecision and bias of a test on its clinical specificity. He defined an analytical "error budget" as the squared sums of the imprecision and bias errors. The maximum limit for this error budget was set at a value corresponding to a 50% increase in the false-positive rate for classifying healthy subjects. For Gaussian distributions with $\pm 2 s$ used as decision limits, this budget equates to $0.45 s_c$. He recommended that this budget be allocated as $0.36 s_c$ for bias, and $0.18 s_c$ for the analytical standard deviation (these fractions do not add up to 0.45 because bias and imprecision are not additive; the bias has a greater influence on false classification than imprecision).

A striking feature is the fact that all of the individual approaches described above recommend numbers for analytical standard deviation near or equal to 0.5 times the biological standard deviation. In addition, the later studies split this number into portions for imprecision and bias, if both are present (35, 51, 55, 56); only Ross & Fraser (54) use the sum of the numbers derived for maximum bias and imprecision, if both are present.

Modern Concepts for Assessing Method Bias

As shown above, only the newer concepts relating analytical performance to medical needs discuss the influence of bias on the medical usefulness of the analytical result. The reason for the late introduction of the bias component in the discussions on desirable performance standards in laboratory medicine is that for a long time only a limited accuracy base existed. This changed, when so called "Definitive Methods" were developed, which are the "best approximation to the true value" of a quality. Based on this approach, a comprehensive measurement system in laboratory medicine was proposed by Tietz (57). An important role in this concept is played by the "Reference Methods", which are intended to transfer the accuracy base of the Definitive Methods to the routine methods (58–60). The bias of methods is of particular importance when intermethod or interlaboratory comparability is required. Both features are primarily assessed in external quality assessment (EQA) surveys and therefore the approach of Tietz was recommended early in this field (9, 61–63). Comparison of routine methods in laboratory medicine with accuracy based methods have been published (64–66) and meanwhile some countries use an accuracy base for performance evaluation for a variety of quantities in their internal and external quality assessment programmes (10, 67, 68).

Comparison of the Different Approaches

In table 1, some examples of the maximum medically useful analytical coefficients of variation proposed in the literature for single individual testing are compared with

Tab. 1 Maximum medically useful analytical coefficients of variation proposed in the literature for single individual testing and analytical "state of the art".

Quantity	Proposed analytical coefficient of variation from references*					
	(69)	(30)	(10)	(9)	(20) ^a	(27)
S-Sodium	0.9	0.3	2.0	1.1	1.7	0.9
S-Chloride	1.2	0.7	2.0	1.5	—	1.0
S-Calcium	1.5	0.9	3.3	1.5	4.8	—
S-Protein	1.5	1.4	3.0	3.3	8.3	3.5
S-Glucose	2.1	2.2	5.0	3.0	11.2	5.0
S-Creatinine	2.0	2.2	6.0	6.7	10.1	—
S-Potassium	1.4	2.4	2.7	5.0	4.8	—
S-Cholesterol	2.6	2.7	6.0	6.0	12.3	5.0
S-Phosphate	2.2	4.0	5.0	5.0	14.3	5.0
S-Uric acid	2.2	4.2	6.0	6.3	—	—
S-Urea	2.5	6.3	8.0	10.0	12.2	5.0

* Median CV_a obtained from the Netherlands external/internal quality assessment scheme (69); Aspen conference (30); German Guidelines (10); Gilbert (9); Skendzel et al. (20), "whenever two values are given, the lower is listed; Tonks (27).

the state of the art as derived from external quality assessment data (69). The concepts related to these examples were discussed earlier in the review part. In all cases the between day analytical coefficient of variation is considered. The data in columns 4 + 5 refer to medical decision limits as specified in the respective publications, while the data in columns 1–3 were given by the respective authors without restriction to specific concentrations. In this way, the decision levels are included in the latter, which allows the comparison of both types of data. The data of *Tonks* (27) in column 6 are shown only for historical reasons. With the exception of S-urea, the goals derived from the within-subject biological variation are the most stringent (column 2). The most liberal goals are those obtained from physicians' interviews (column 5). Between them lie the goals advocated by experts and national recommendations (columns 3 + 4). Interestingly, most of these follow in general the demands given by the biological variation of the quantities, e. g. for S-sodium the proposed CV are always the lowest ones. For S-uric acid or S-urea usually quite high CV are tolerated. But there are examples where quantities with comparable biological variations show pronounced differences, e. g. S-creatinine and S-potassium. The values for S-potassium are considerably lower (see columns 3–5). This could be due to the fact that many methods for S-creatinine today still have accuracy and precision problems, while S-potassium normally is measured with methods of higher precision. It seems that the goals for S-creatinine and S-potassium derived by experts, external quality assessment organizers or questionnaires of clinicians reflect more or less the analytical state of the art for these two analytes. The comparison of the state of the art CV (69) (column 1) with the most stringent goals proposed at the *Aspen Conference* (30) (column 2) shows that the analytical precision is now satisfactory for most of the presented quantities, provided the methods have no bias. Exceptions are analytes such as S-sodium, S-chloride and S-calcium, for which the biological variation is very small. An interesting feature in the numbers advocated by *Tonks* (27) (column 6) is the tendency to set an upper imprecision limit, even if biology would allow higher values. This might reflect the fact that different quantities are often measured with similar methods and equipment, which results in similar imprecision.

Comments of the Working Group on the Approaches in the Literature for Deriving Analytical Quality Specification from:

(i) The state of the art

Advantages of this concept are that analytical quality specifications can easily be established and most

laboratories will accept them. Disadvantages are that they might not be related to optimal patient care and that these analytical quality specifications change with time. The state of the art can improve, but it can also deteriorate.

(ii) The views of experts and expert groups

Advantages are that they can be easily established on a local basis and will gain a high acceptance in laboratories. Disadvantages are that these analytical quality specifications often cannot be transferred to a broader community and that they often are influenced by the analytical state of the art.

(iii) Opinions of clinicians

The advantages of this concept are that it may be related to the clinicians' perception when to react. In addition, the analytical quality specifications derived represent more or less the situation used in practice. Disadvantages are that often different opinions among clinicians exist and that it is often impossible to fix the probability level for significance of results. In addition, the analytical quality specifications derived might be very much influenced by the analytical state of the art. The latter will be discussed in more detail below. Some studies made an attempt to derive analytical quality specifications from the viewpoint of clinicians by sending questionnaires to experienced physicians in many different fields of laboratory medicine (17–22). They were asked for decision levels for common quantities in particular clinical situations or for changes in concentrations which would cause their medical action. It was shown that the decision limits given in the work of *Skendzel et al.* (20) will cause < 5% false positive alarms for most of the common quantities, when the observed analytical results exceed these action limits, provided the results were obtained with state of the art analytical performance (13). This supports the conclusion that these action limits might be influenced to a great extent by the analytical state of the art. But it was also stated that the medical significant differences may be too narrow for some quantities, because the clinicians seem to underestimate the biological variation of those quantities (13). In addition, it can be questioned, whether others are not unnecessarily high. To investigate this problem in more detail, the within-subject biological variation (CV_i), the analytical variation (CV_a), and the medical decision values must be viewed in relation to each other. Table 2 compares the quotients of CV_a and CV_i , Δ_{med} and CV_i and CV_i for some of the common quantities [they are arranged by descending CV_a/CV_i values, and the data of *Linnet* (13) are used]. From these calculations it is

Tab. 2 Comparison of CV_a/CV_i , Δ_{med}/CV_i , and CV_i of 17 common quantities arranged by descending CV_a/CV_i values [the data of l.c. (13) are used for the calculation].

Quantity	CV_a/CV_i	Δ_{med}/CV_i	CV_i
S-Sodium	1.49	4.5	0.8
S-Creatinine	1.39	5.89	4.4
S-Calcium	1.33	5.39	1.8
S-Thyroxine	1.18	5.0	7.6
Prothrombin time	0.85	5.58	4.5
S-Glucose	0.79	5.09	4.4
S-Cholesterol	0.79	4.23	4.8
S-Total protein	0.74	4.57	3.0
S-Phosphate	0.60	4.43	5.8
S-Potassium	0.48	2.70	4.4
S-Bilirubin	0.48	1.87	23
S-Aspartate aminotransferase	0.43	2.39	14
S-Urea	0.37	2.48	12.4
Haemoglobin	0.24	1.69	4.5
S-Triacylglycerols	0.21	1.02	26
Leukocyte count	0.17	1.71	14.7
S-Iron	0.12	1.09	26

surprising that the quantities with a low (< 0.5) CV_a/CV_i value (which indicates sufficient analytical precision) also have a low Δ_{med}/CV_i value (the “a priori” least medical significant difference which analytically can be detected for two consecutive measurements ($P = 0.05$) is: $\Delta_{med} = 2.77 \times CV_i$, which would result in an ideal Δ_{med}/CV_i value of 2.77). It seems that the medical decision limits have been chosen to be quite narrow despite the fact that these quantities mostly have a great biological variation. In other words, this variation seems to be underestimated by clinicians with the consequence that the number of false positive results is high (13). On the other hand, many quantities with a high (> 0.5) CV_a/CV_i value (which indicates insufficient analytical precision) have high Δ_{med}/CV_i values. In these cases, it can be questioned whether the medical decision limits should not be lowered. Maybe the poor analytical imprecision has been overestimated here by the clinicians. A result of this is a relatively high number of false negative results. Therefore, analytical goal setting on the basis of the viewpoint of clinicians seems to be more influenced by “the state of the art” than by what is desirable. In addition, it may lead to incongruent decision limits when different quantities are compared.

(iv) Analytical goals and the needs of quality control

As stated above, this approach usually uses only currently defined goals and investigates whether they are in accordance with the needs for quality control.

(v) Analytical goals derived from specific clinical situations

Advantages of this approach are that generally analytical imprecision and bias are evaluated. In addition, the goals derived are directly related to the clinical use. They are most useful when one particular test is used for one specific clinical situation. Disadvantages are that they are complicated, often require much effort for their production, mostly can be used only on a local basis, and that the clinical strategy must be clearly defined. In addition, this approach lacks practicability for laboratory tests that are used for many different clinical situations. Nevertheless, the group advocates this approach wherever it can be applied. A detailed description of its application in various clinical strategies can be found in *Hyltoft Petersen & Hørdér* (1).

(vi) Analytical goals based on within- and between-subject biological variation

The basis of this concept is that the mean within-subject biological variation is nearly constant in all investigations (49) including the elderly (70) and groups of patients in certain stable states (49). This allows the transfer of data over time and geography. Further, where the population is homogeneous for a quantity, common reference intervals can be used within large regions or within ethnic groups. This can save considerable resources in laboratories which otherwise have to establish their own reference intervals. In addition, the concept of common “reference changes” (33) could be applied in the whole of Europe (and even more widely). For a more recent article on the concept of reference changes see *Queralto et al.* (50).

The Concept of the Working Group

Clarification of the terms “goal” and “desirable analytical performance standard”

Sometimes confusion arises over the use of the word “goal” for both medical and analytical aims; in fact they should be clearly separated. A medical goal is, e.g. the use of common reference intervals. For this medical goal, a “desirable analytical performance standard” (= “analytical goal”) must be derived. This desirable analytical performance standard must then be specified in terms of maximum method bias and imprecision. When achieved, these performance standards should guarantee optimal patient care. To illustrate this way of thinking, Scheme 1 presents some analytical goals for different application levels (we use the term “analytical goal” in the following always in the sense of “desirable performance standards”, as described above). Two main levels

Scheme 1 Goals and related desirable analytical performance standards at different levels

Level	Goal	Desirable performance standards
Achieving medical utility	1. Make use of:	
	a. Common reference changes	In the absence of systematic changes: $s_a \leq 0.5 s_i$ (see Annex for detailed discussion)
	b. Common reference intervals	When imprecision is negligible: $B \leq 0.25 s_c$ (see Annex for detailed discussion)
	2. Analytical performance should not influence the outcome of clinical strategies	The demands implied for desirable performance should be evaluated for each clinical strategy according to accepted models for the specific purpose (diagnosis, screening, monitoring, etc.)
Accreditation	Guard against unprofessional performance	The specifications must consider the analytical state of the art, but should gradually be converted in accordance with the above recommendations.

are shown in Scheme 1, which indicate differences in approach. Achieving medical utility is always related to biology whereas accreditation systems are mostly related to the actual analytical performance attained in laboratories. The latter uses goals from biological variation only if convenient. This is practical, since it reflects what is possible with the methods and equipment in use, but it may not reflect the medical utility. It is obvious that also the EQA of accreditation/licensing testing has to consider the state of the art performance, but a gradual conversion of systems should be undertaken in accordance with medical utility concepts whenever possible.

Recommendations of the Working Group

From the above discussions, the Working Group recommends that analytical quality specifications be based on biology as a realistic basis, and that the approaches (i) to (iii) are reasonable only for providing provisional solutions: the first in assessing surveys and the other two as interesting approaches in individual laboratories. The starting point for our work was, therefore, the concept of *Harris* (32) and *Cotlove et al.* (28) for patient monitoring: $s_a \leq 0.5 s_i$, and the approach of *Gowans et al.* (35) for diagnostic testing: $B \leq 0.25 s_c$. We expanded the concept of *Cotlove et al.* (28) because, during monitoring, unidirectional systematic changes (ΔSE) (= drifts) can occur, and we accepted the concept of *Gowans et al.* (35), who based their goals on the recommendations of the IFCC for the establishment of reference intervals (52). This concept is especially useful for diagnostic testing. A detailed discussion of the derivation of analytical quality specifications for monitoring and diagnostic testing can be found in the Annex. In particular, the combined effect of bias and imprecision is presented in a graphical form. In short, we recommend that desirable performance standards should be calculated

for monitoring as:

$s_a \leq 0.5 s_i$
(in the absence of unidirectional systematic changes), or

$\Delta SE \leq 0.33 s_i$ (when imprecision is negligible);
see also Annex;

for diagnostic testing as:

$B \leq 0.25 s_c$ (when the imprecision is negligible),
or
 $s_a \leq 0.58 s_c$ (when bias is negligible);
see also Annex.

Data on biological variation for most of the common serum quantities are available (49, 71–73) and the desired performance standards can be calculated as shown in table 3 (74). If data for s_g are not available, an approximation for bias can be 1/16 of the range of the reference

Tab. 3 Specifications for desirable analytical performance (from l. c. (74))

Quantity	Maximum CV_a in the absence of syst. changes (using $CV_a \leq 0.5 CV_i$)	Maximum bias (%) if the imprecision is negligible (using $B \leq 0.25 CV_c$)
S-Sodium	0.3	0.2
S-Calcium	0.9	0.7
S-Creatinine	2.2	2.8
S-IgM	2.3	12.7
S-Potassium	2.4	1.6
S-Transferrin	2.4	2.3
S-Cholesterol	2.7	4.1
S-Lactate	3.9	4.1
S-Phosphate	4.0	3.1
S-Bilirubin	11.3	9.8
S-Triacylglycerols	11.6	15.6
S-Alanine aminotransferase	13.6	13.6

interval (this assumes that s_a is small in relation to s_i and s_g). We recommend that these desirable performance standards should also be introduced in external quality assessment as the basis for calculation of acceptance limits. When more demanding specifications based on documented medical needs are stated, these may be substituted for those based on biological variation. The goals outlined above may be viewed as aims for the future if they cannot be reached with current analytical equipment and methodology. Last, but not least, the group recommends that limits for interference and unspecificity should be related to the biological variation of the respective quantity (75). This is more relevant than the fixed limits used previously [e.g. 5% for electrolytes and substrates and 10% for enzymes (76); or 3% (77), and 10% (78) for all quantities]. Further, the group recommends the use of goals derived from specific clinical situations (1, 79) whenever they are more appropriate.

Comparison of the Recommendations with the Actual Situation

It should be noted that, for a number of quantities, the "state of the art" analytical precision is better than the goals. Consider for example S-triacylglycerols. The proposed maximum coefficient of variation was 11.6% (see tab. 3). If unidirectional systematic changes are also taken into account, it must be reduced to maybe 6 or 8%. Much lower values are attainable today. Therefore, in the opinion of the Working Group, it would neither be wise nor reasonable if a new triacylglycerol method with a CV in the range of 6% was introduced on the market. First, it must not be forgotten that an analytical standard deviation of $0.5 s_i$, still adds 12% variation to the true test variation. In addition, it generally can be stated that the better a method performs, the less time and money has to be invested in its control (14). Therefore, in the above cited example, more stringent criteria than those derived from the biological variation are advisable. On the other hand, there are many quantities whose analytical precision is insufficient according to the above cited concept, especially when method bias is taken into account. Insufficient analytical precision is normally associated with quantities which have narrow biological variation like S-sodium, S-chloride, S-calcium or S-protein, despite the attainment of quite low CV values for these quantities. Therefore, in these cases, performing more than one measurement might be more effective than aiming for a CV below 0.5%. Then less stringent criteria than those derived above could be accepted. Even more difficult to achieve is a method bias below 0.5%, because then the temperature during analysis must be kept constant or be constantly

monitored, because densities of solutions, and therefore their concentrations, change with temperature. This has an effect of about 0.12% within a temperature change of 5 degrees. In addition buoyancy corrections should be made for weighings, and so on (80). Another consequence would be the necessity to keep preanalytical errors below 0.5%, otherwise the final result would be dominated by the preanalytical errors. Taking all these sources of error together, routine method inaccuracies below 1% are extremely difficult to achieve, it may even not be possible to assess them and they might be outweighed by the preanalytical errors. Therefore, accuracy seems to be a major challenge for future developments.

Conclusion

The proposed "desirable performance standards" are based on within- and between-subject biological variation for several reasons, namely

- a) the model is simple to understand and apply,
- b) there are many data on biological variation,
- c) the within- and between-subject biological variations are nearly constant over geography, and
- d) when these criteria are fulfilled, the analytical quality will satisfy most clinical needs.

Furthermore, it is often difficult to evaluate clinical needs in complicated clinical situations in which other considerations may be more relevant in the process of decision-making. However, when more demanding clinically derived specifications are stated, these might be substituted for those based on biological variation. This may be the case for S-cholesterol (tab. 3) for which the National Cholesterol Education Program (11) recommended a bias of less than 3%. Further, we recommend that where the state of the art is much better than the desirable performance standards derived from the biological variation, this should be welcomed because it gives the user additional benefits, especially in the internal quality control of such methods.

ANNEX

Detailed Concept of the Working Group for Deriving Analytical Quality Specifications from Biology

Abbreviations used:

Within-subject (individual) biological standard deviation:

s_i

Between-subject (group) biological standard deviation:

$$s_g$$

Composite biological standard deviation:

$$s_c = (s_i^2 + s_g^2)^{1/2}$$

Analytical imprecision expressed as standard deviation:

$$s_a$$

Coefficient of variation:

CV (the same subscripts as above are used, e. g. CV_i)

Bias: B

Drift (unidirectional changes in bias):

$$\Delta SE$$

Shift (difference in bias when two methods are used for monitoring the same patient, see l. c. (81)):

$$\Delta SE$$

Reference interval (for normal distributed quantity):

$$R (= \mu \pm 1.96 s)$$

Medically significant difference (reference change):

$$\Delta_{med}$$

General assumptions underlying the concepts

We wanted to develop a concept to be generally applicable in the routine laboratory. Consider, for example, the case of a potassium test to be used in a general medical laboratory. The test should cover potassium determination in plasma in emergency situations, such as digitalis poisoning, heart arrhythmias, acute renal failure, or diarrhoea; in therapy decisions, such as treatment with angiotensin-converting enzyme blocker, digitalis, or diuretics; or in monitoring illness, such as diabetes mellitus, chronic renal failure, or gastrointestinal diseases. These are situations, which today, can normally all be covered by one and the same test. Developing general goals for such a test means that the proposed goals should be independent of the concentration of potassium present in a specimen. It also follows that we always considered a *Gaussian* distribution and 5% test-level. Therefore, these goals are not ideal because they are not the best choice for each particular application. But, on the other hand, it is our conviction that they are especially useful for the daily routine operation of a typical medical laboratory. We advocate to use % bias and CV to express the maximum analytical bias and imprecision derived by the concept below, because laboratories are more familiar with this way of presentation than with absolute numbers of bias or standard deviations (for examples see tab. 3).

Analytical quality specifications for monitoring

The analytical specifications shall be derived from the least "medically significant difference" ("reference

change") concept (33). This concept considers the case of two consecutive measurements (x_1 and x_2) performed under identical analytical conditions, with the assumption of normal distributed x values. In addition both measurements are performed on the same underlying population ($\mu_1 = \mu_2$). Then, the smallest medically significant difference (Δ_{med}) which analytically can be detected for two consecutive measurements ($P = 0.05$) ($s_a = 0$) is:

$$\Delta_{med} = 1.96 \times \sqrt{2} \times s_i = 2.77 \times s_i \text{ (i).}$$

Two important cases have to be considered for monitoring: a) $\Delta SE > 0$; b) $s_a > 0$. Therefore, equation (i) has to be rewritten as follows (25):

$$\Delta_{med} = 2.77 \times (s_a^2 + s_i^2)^{1/2} + \Delta SE \text{ (ii).}$$

If $\Delta SE = 0$, Δ_{med} is only dependent on s_a and s_i , and we can use the concept of *Harris* $s_a \leq 0.5 s_i$ (32). From this follows:

$$\begin{aligned} \Delta_{med} &\leq 2.77 (0.25 s_i^2 + s_i^2)^{1/2} \\ \Delta_{med} &\leq 3.10 s_i \text{ (iii).} \end{aligned}$$

Now the case $s_a = 0$ shall be considered. We start again with equation (ii); with $s_a = 0$, and $\Delta_{med} \leq 3.10 s_i$ as calculated in (iii):

$$\begin{aligned} 2.77 \times s_i + \Delta SE &\leq 3.10 s_i, \text{ giving} \\ \Delta SE &\leq 0.33 s_i \text{ (iv).} \end{aligned}$$

At this stage, we have derived maximum values for s_a (with $\Delta SE = 0$) and ΔSE (with $s_a = 0$). Now the function describing the combined effect of both is presented. The combined effects of drift and analytical imprecision, related to biology, can be represented by figure 1 (25). It covers the case for which, due to analytical uncertainty, the ideal medically significant difference of $2.77 s_i$ has to be widened to $3.10 s_i$. It should be noted

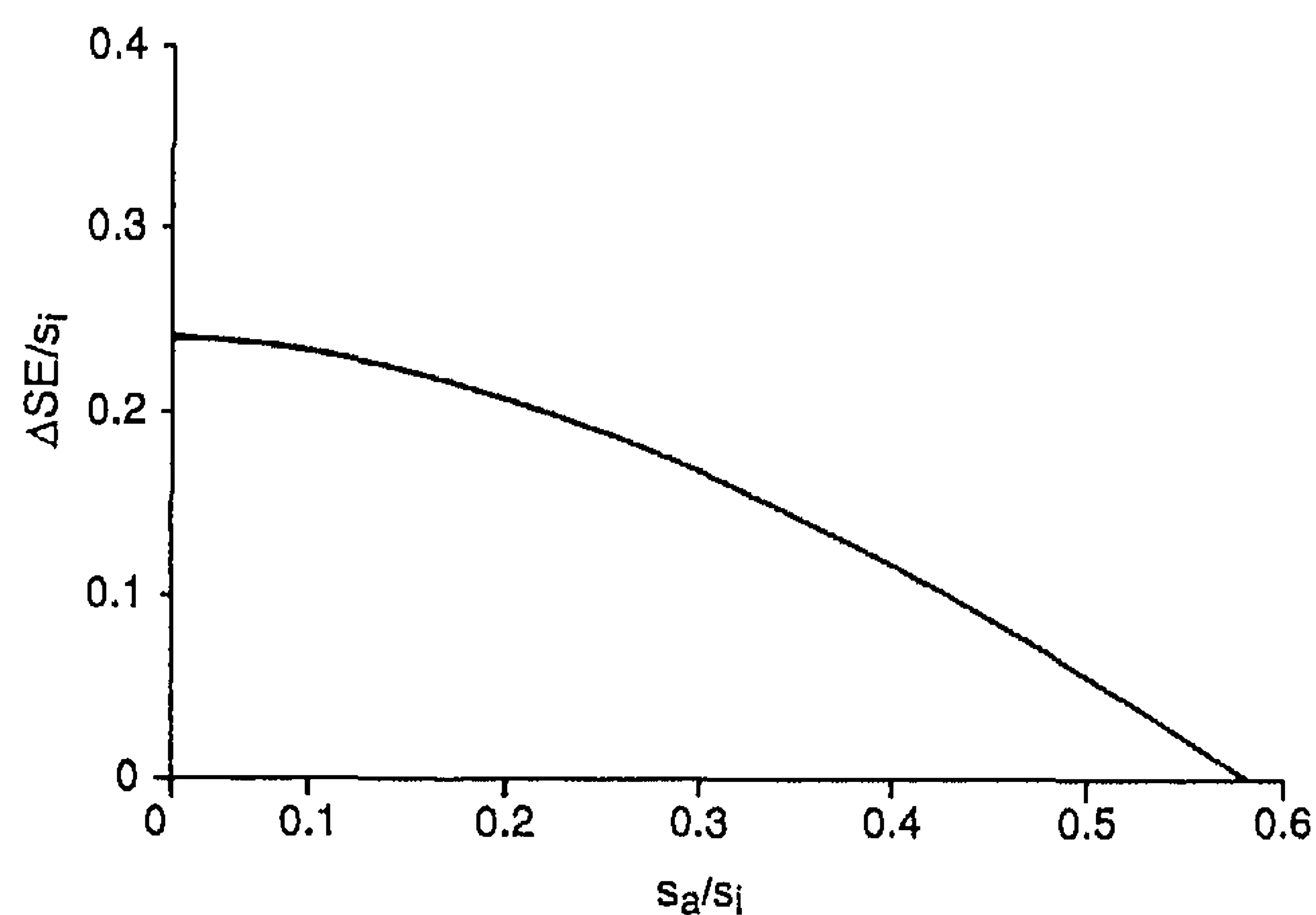


Fig. 1 Reference change (Δ_{med}) concept: relationship between maximum allowable drift and analytical standard deviation in fraction of s_i , ($\Delta SE/s_i$) and (s_a/s_i), respectively. The figure represents the case $\Delta_{med} = 3.10 s_i$ [from l. c. (25), with permission].

that any combination of bias and standard deviation below this curve would allow reduction of Δ_{med} to values between 2.77 and 3.10 s_i . Further, the values of the function for $\Delta \text{SE} = 0$ and $s_a = 0$ are the same as calculated above.

Analytical quality specifications for diagnostic testing

The analytical specifications are outlined for sharing common reference intervals, according to the concept of *Gowans et al. (35)*. This concept investigates the influence of method bias and imprecision on the percentage of individuals outside each reference limit for a normal distributed reference population. In figure 2 (35), the combined effects of bias and imprecision place 4.6% of a population outside of each reference limit (with zero imprecision and bias, this is 2.5%). Again, the two extreme cases, bias = 0 or $s_a = 0$ can be read from figure 2:

$$s_a \leq 0.58 s_c \text{ (v), for bias} = 0$$

$$B \leq 0.25 s_c \text{ (vi), for imprecision} = 0.$$

References

1. Hyltoft Petersen P, Hørdér M. Influence of analytical quality on test results. *Scand J Clin Lab Invest* 1992; 52 Suppl 208:65–87.
2. Fraser CG, Hyltoft Petersen P. Desirable standards for laboratory tests if they are to fulfill medical needs. *Clin Chem* 1993; 39:1447–55.
3. Linnet K. A review on the methodology for assessing diagnostic tests. *Clin Chem* 1988; 34:1379–86.
4. Büttner J. Laboratory findings: structure, validity and significance for medical cognitive processes. *Eur J Clin Chem Clin Biochem* 1991; 29:507–19.
5. Christensen M. Clinical test evaluation. Basic concepts. *Scand J Clin Lab Invest* 1992; 52 Suppl 208:13–29.
6. Henderson AR. Assessing test accuracy and its clinical consequences: a primer for receiver operating characteristic curve analysis. *Ann Clin Biochem* 1993; 30:521–39.
7. Barnett RN. Error limits and quality control [editorial]. *Arch Pathol Lab Med* 1989; 113:829–30.
8. Barnett RN. Medical significance of laboratory results. *Am J Clin Pathol* 1968; 50:671–6.
9. Gilbert RK. Progress and analytical goals in clinical chemistry. *Am J Clin Pathol* 1975; 63:960–73.
10. Richtlinien der Bundesärztekammer zur Qualitätssicherung in medizinischen Laboratorien. *Dt Ärztebl* 1988; 85:A699–A712.
11. Naito HK. Current status of blood cholesterol measurement in clinical laboratories of the United States: a report from the Laboratory Standardization Panel of the National Cholesterol Education Program. *Clin Chem* 1988; 34:193–201.
12. Campbell BG. Evaluation of two types of "medically significant error limits" and two quality control procedures on a multichannel analyzer. *Arch Pathol Lab Med* 1989; 113:834–7.
13. Linnet K. Choosing quality control systems to detect maximum clinically allowable analytical errors. *Clin Chem* 1989; 35:284–8.
14. Westgard JO, Burnett RW. Precision requirements for cost-effective operation of analytical processes. *Clin Chem* 1990; 36:1629–32.
15. Westgard JO, Petersen PH, Wiebe DA. Laboratory process specifications for assuring quality in the U. S. National Cholesterol Education Program. *Clin Chem* 1991; 37:656–61.
16. Bennett ST, Connelly DP, Eckfeldt JH. Assessment of split-sample proficiency testing for cholesterol by use of a computer simulation model. *Clin Chem* 1991; 37:497–503.
17. Campbell DG, Owen JA. The physician's view of laboratory performance. *Aust Ann Med* 1969; 18:4–6.
18. Barrett AE, Cameron SJ, Fraser CG, Penberthy LA, Shand KL. A clinical view of analytical goals in clinical biochemistry. *J Clin Pathol* 1979; 32:893–6.
19. Elion Gerritzen WE. Analytical precision in clinical chemistry and medical decisions. *Am J Clin Pathol* 1980; 73:183–95.
20. Skendzel LP, Barnett RN, Platt R. Medical useful criteria for analyte performance of laboratory tests. *Am J Clin Pathol* 1985; 83:200–5.
21. Lott JA, Surufka N, Massion CG. Proficiency testing of serum enzymes based on medical needs criteria. *Arch Pathol Lab Med* 1991; 115:11–4.
22. Thue G, Sandberg S, Fugelli P. Clinical assessment of haemoglobin values by general practitioners related to analytical and biological variation. *Scand J Clin Lab Invest* 1991; 51:453–9.
23. Petersen PH, Hørdér M. Ways of assessing quality goals for diagnostic tests in clinical situations. *Arch Pathol Lab Med* 1988; 112:435–43.
24. Fraser CG, Petersen PH, Larsen ML. Setting analytical goals for random analytical error in specific clinical monitoring situations. *Clin Chem* 1990; 36:1625–8.
25. Larsen ML, Fraser CG, Petersen PH. A comparison of analytical goals for haemoglobin A_{1c} assays derived using different strategies. *Ann Clin Biochem* 1991; 28:272–8.

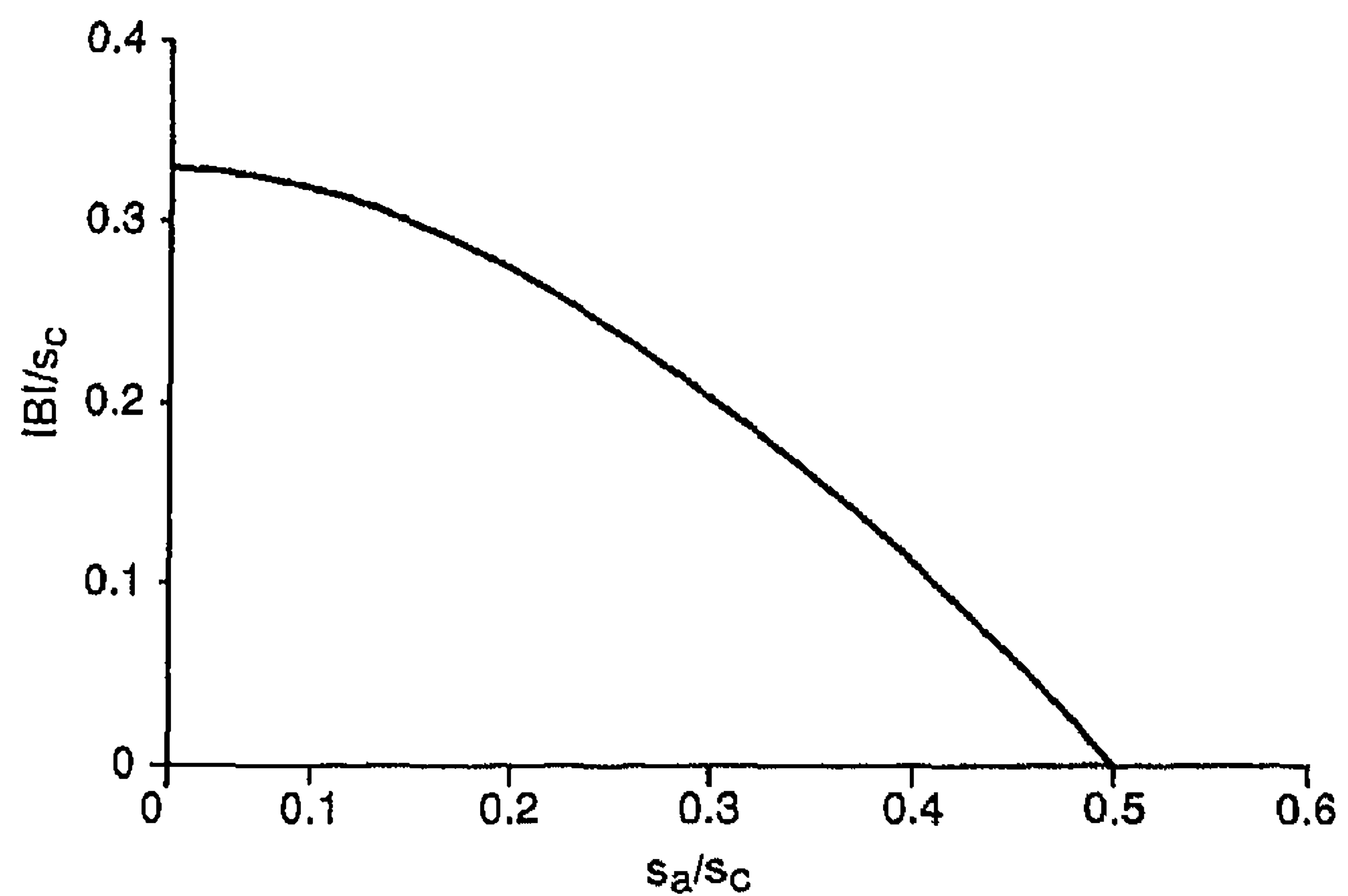


Fig. 2 Common reference interval concept: relationship between maximum allowable bias and analytical standard deviation in fraction of s_c , ($|B|/s_c$) and (s_a/s_c), respectively. The figure represents the case that 4.6% of the individuals are outside each reference limit due to analytical uncertainty [from l. c. (35), with permission].

Acknowledgement

We are indebted to A. Uldall for continuous help for meeting facilities and stimulation of the project. We thank all members of the Working Groups for fruitful discussions during our common meetings and for exchange of ideas by numerous letters.

26. Jenny RW. Analytical goals for determination of theophylline concentration in serum. *Clin Chem* 1991; 37:154-8.
27. Tonks DB. A study of accuracy and precision of clinical chemistry determinations in 170 Canadian laboratories. *Clin Chem* 1963; 9:217-33.
28. Cotlove E, Harris EK, Williams GZ. Components of variation in long term studies of serum constituents in normal subjects. III. Physiological and medical implications. *Clin Chem* 1970; 16:1028-32.
29. Glick JH Jr. Expression of random analytical error as a percentage of the range of clinical interest. *Clin Chem* 1976; 22:475-83.
30. Elevitch FR, editor: *Analytical goals in clinical chemistry*. Proceedings of the CAP Aspen Conference Aug 25-27 1976, Skokie (IL). Skokie (IL): College of American Pathologists, 1977.
31. Analytical goals in clinical chemistry: Their relationship to medical care. Proceedings of the Subcommittee on Analytical Goals in Clinical Chemistry, World Association of Societies of Pathology, CIBA Foundation; 1978 Apr 25-28; London. *Am J Clin Pathol* 1979; 71:624-30.
32. Harris EK. Statistical principles underlying analytical goal-setting in clinical chemistry. *Am J Clin Pathol* 1979; 72:374-82.
33. Harris EK, Yasaka T. On the calculation of a "reference change" for comparing two consecutive measurements. *Clin Chem* 1983; 29:25-30.
34. Tietz NW. *Textbook of clinical chemistry*. Philadelphia: Saunders, 1986.
35. Gowans EMS, Hyltoft Petersen P, Blaabjerg O, Hørdler M. Analytical goals for the acceptance of common reference intervals for laboratories throughout a geographical area. *Scand J Clin Lab Invest* 1988; 48:757-64.
36. Hyltoft Petersen P, Gowans EMS, Blaabjerg O, Hørdler M. Analytical goals for the estimation of non-Gaussian reference intervals. *Scand J Clin Lab Invest* 1989; 49:727-37.
37. Fraser CG, Harris EK. Generation and application of data on biological variation in clinical chemistry. *Crit Rev Clin Lab Sci* 1989; 27:409-37.
38. Department of Health and Human Services Health Care Financing Administration. Medicare, Medicaid and CLIA programs; regulations implementing the Clinical Laboratory Improvement Amendments of 1988 (CLIA). *Fed Regist* 1992; 57:7002-186.
39. Büttner J, Borth R, Boutwell JH, Broughton PMG, Bowyer RC. Approved recommendations on quality control in clinical chemistry. Part 6. Quality requirements from the point of health care. *J Clin Chem Clin Biochem* 1980; 18:861-6.
40. Büttner J, Borth R, Boutwell JH, Broughton PMG, Bowyer RC. Approved recommendations on quality control in clinical chemistry. Part 5. External quality control. *J Clin Chem Clin Biochem* 1983; 21:885-92.
41. de Verdier CH, Aronsson T, Nyberg A. Quality control in clinical chemistry. Efforts to find an efficient strategy. *Scand J Clin Lab Invest* 1984; 44 Suppl 72:9-241.
42. Westgard JO, Carey RN, Wold S. Criteria for judging precision and accuracy in method development and evaluation. *Clin Chem* 1974; 20:825-33.
43. Westgard JO, Hunt MR. Use and interpretation of common statistical tests in method comparison studies. *Clin Chem* 1973; 19:49-57.
44. Westgard JO, Groth T. Power functions for statistical control rules. *Clin Chem* 1979; 25:863-9.
45. Fraser CG. Better criteria for desirable laboratory performance exist [letter]. *Am J Clin Pathol* 1986; 85:251.
46. Schectman G, Sasse E. Variability of lipid measurements: relevance for the clinician. *Clin Chem* 1993; 39:1495-503.
47. Lott JA, Manning NR, Kyler MK. Proficiency testing in a medical needs context. *Clin Chem* 1989; 35:347-54.
48. Gowenlock AH, Broughton PMG. The influence of accuracy and precision on the normal range and on acceptable limits for an analytical result. *Z Anal Chem* 1968; 243:774-80.
49. Fraser CG. The application of theoretical goals based on biological variation data in proficiency testing. *Arch Pathol Lab Med* 1988; 112:404-15.
50. Querafio JM, Boyd JC, Harris EK. On the calculation of reference change values, with examples from a long-term study. *Clin Chem* 1993; 39:1398-1403.
51. Harris EK. Proposed goals for analytical precision and accuracy in single point testing. *Arch Pathol Lab Med* 1988; 112:416-20.
52. Solberg HE, Petit Clerc C. Approved recommendation on the theory of reference values. Part 3. Preparation of individuals and collection of specimens for the production of reference values. *J Clin Chem Clin Biochem* 1988; 26:593-8.
53. Ross JW. A theoretical basis for clinically relevant proficiency testing evaluation limits: sensitivity analysis of the effect of inherent test variability on acceptable method error. *Arch Pathol Lab Med* 1988; 112:421-34.
54. Ross JW, Fraser MD. Analytical goals developed from inherent error of medical tests. *Clin Chem* 1993; 39:1481-94.
55. Stamm D. Die neuen Richtlinien der Bundesärztekammer zur Qualitätssicherung in medizinischen Laboratorien. *DG Klin Chem Mitteil* 1990; 21:69-77.
56. Klee GG. Tolerance limits for short-term analytical bias and analytical imprecision derived from clinical assay specificity. *Clin Chem* 1993; 39:1514-8.
57. Tietz NW. A model for a comprehensive measurement system in clinical chemistry. *Clin Chem* 1979; 25:833-9.
58. Boutwell JH, editor: *A national understanding for the development of reference materials and methods for clinical chemistry*. Proceedings of a conference. Washington: American Association for Clinical Chemistry, 1978.
59. Büttner J. Reference methods as a basis for accurate measuring systems. *Eur J Clin Chem Clin Biochem* 1991; 29:223-35.
60. Stöckl D, Reinauer H. Development of criteria for the evaluation of reference method values. *Scand J Clin Lab Invest* 1993; 53 Suppl 212:16-8.
61. Gilbert RK. CAP interlaboratory survey data and analytical goals. Elevitch FR, editor: *Analytical goals in clinical chemistry*. Proceedings of the CAP Aspen Conference Aug 25-27 1976, Skokie (IL). Skokie (IL): College of American Pathologists, 1977:63-73.
62. Stamm D. Reference materials and reference methods in clinical chemistry. *J Clin Chem Clin Biochem* 1979; 17:283-97.
63. Stamm D. A new concept for quality control of clinical laboratory investigations in the light of clinical requirements and based on reference method values. *J Clin Chem Clin Biochem* 1982; 20:817-24.
64. Gilbert RK. Accuracy of clinical laboratories studied by comparison with definitive methods. *Am J Clin Pathol* 1978; 70 Supplement without no.: 450-70.
65. Hartmann AE, Naito HK, Burnett RW, Welch MJ. Accuracy of participant results utilized as target values in the CAP chemistry survey program. *Arch Pathol Lab Med* 1985; 109:894-903.
66. Stöckl D, Reinauer H, Thienpont LM, De Leenheer AP. Assessment of the accuracy of routine assays for the determination of steroid hormones and thyroxine in serum against isotope dilution mass-spectrometry target values. *Comm Lab Med* 1992; 1:77-9.
67. Welch MJ, Hertz HS. The how and why of an accuracy base for proficiency testing programs. *Arch Pathol Lab Med* 1988; 112:343-6.
68. Ehrmeyer SS, Laessig RH, Leinweber JE, Oryall JJ. 1990 Medicare/CLIA final results for proficiency testing: maximum intralaboratory performance characteristics (CV and bias) needed to pass. *Clin Chem* 1990; 36:1736-40.
69. Jansen RTP, Bullock DG, Vassault A, Baadenhuijsen H, De Leenheer A, Dumont G, de Verdier C-H, Zender R. Between-country comparability of clinical chemical results: an international quality assessment survey of 17 analytes in six European countries through existing national schemes. *Ann Clin Biochem* 1993; 30:304-14.

70. Fraser CG, Wilkinson SP, Neville RG, Knox JDE, King JF, MacWalter RS. Biologic variation of common hematologic quantities in the elderly. *Am J Clin Pathol* 1989; 92:465–71.
71. Libeer JC. External Quality Assessment in clinical laboratories, European perspectives, today and tomorrow [dissertation]. Antwerp, Belgium: Univ. of Antwerp, 1993.
72. Ricos C, Arbos MA. Quality goals for hormone testing. *Ann Clin Biochem* 1990; 27:353–8.
73. Smith SJ, Cooper GR, Myers GL, Sampson EJ. Biological variability in concentrations of serum lipids: sources of variation among results from published studies and composite predicted values. *Clin Chem* 1993; 39:1012–22.
74. Fraser CG, Hyltoft Petersen P, Ricos C, Haecckel R. Proposed quality specifications for the imprecision and inaccuracy of analytical systems for clinical chemistry. *Eur J Clin Chem Clin Biochem* 1992; 30:311–7.
75. Fuentes Arderiu X, Fraser CG. Analytical goals for interference. *Ann Clin Biochem* 1991; 28:393–5.
76. Vassault A, Grafmeyer D, Naudin CL, Dumont G, Bailly M, Henny J, et al. Société Française de Biologie Clinique. Protocole de validation de techniques. *Ann Biol Clin* 1986; 44:686–745.
77. Glick MR, Ryder KW, Jackson SA. Graphical comparison of interferences in clinical chemistry instrumentation. *Clin Chem* 1986; 32:470–5.
78. Ryder KW, Glick MR. Erroneous laboratory results from hemolyzed, icteric, and lipemic specimens. *Clin Chem* 1993; 39:175–6.
79. de Verdier C-H, Groth T, Hyltoft Peterson P, editors. Medical need for quality specifications in clinical laboratories. *Upsala J Med Sci* 1993; 8:187–491.
80. Stöckl D. Desirable performance criteria for quantitative measurements in medical laboratories based on biological analyte variation – hindrances to reaching some and reasons to surpass some. *Clin Chem* 1993; 39:913–4.
81. Hyltoft Petersen P, Fraser CG, Westgard JO, Lytken Larsen M. Analytical goal-setting for monitoring patients when two analytical methods are used. *Clin Chem* 1992; 38:2256–60.

Dr. Dietmar Stöckl
INSTAND e. V.
Johannes-Weyer-Straße 1
D-40225 Düsseldorf
Germany